

Rational Design, Synthesis and Evaluation: Affinity Ligands

G. Gupta and C. R. Lowe,

Institute of Biotechnology, University of Cambridge,
Cambridge, UK

Copyright © 2000 Academic Press

Introduction

For years, the structure, function and dynamics of proteins have intrigued biochemists and chemists and inspired the design of targeted low molecular weight drugs. Today, as biotechnology progresses, proteins are themselves being used in the treatment of several diseases. Almost every gene can now be cloned and its protein product produced in any expression system depending on the desired folding, stability, post-translational modification, cost of production and ease of recovery. An increase in the isolation of large volumes of recombinant proteins such as antibodies, antivirals, cytokines, enzymes, clotting factors and vaccines for therapeutic, diagnostic and research purposes has made efficient protein purification the most important step in the recovery of efficacious and fully active biopharmaceuticals.

All steps in the production of biopharmaceuticals need to be compliant with the safety guidelines laid out by regulatory authorities such as the US Food and Drug Administration and the equivalent European authority. The final product should be a 'well characterized biologic' with defined major impurities, if any, very low levels of DNA and viruses (less than 10 pg per dose), pyrogens (less than 300 endotoxin units per dose) and leachates from the separation matrices. **Table 1** highlights some of the quality control requirements imposed by the regulatory authorities on proteins used for clinical applications. Commonly purified end products are mixtures of isoforms of proteins with variations in post-translational modifications such as glycosylation, oxidation, end terminal alterations, misfolding, incorrect disulfide bridging and nicked or truncated variants. Thus, techniques such as high performance liquid chromatography (HPLC), peptide mapping, capillary electrophoresis, isoelectric focusing, circular dichroism and mass spectrometry are also gaining an impetus in thorough characterization of the potency, purity and safety of protein drugs.

The increasing demand for peptides, nucleotides, low molecular weight synthetic molecules and biomimetic ligands for protein purification has inspired chemists to generate focused and general combinatorial libraries of diverse compounds. Rational ligand

or drug design can provide directionality and help increase the possibility of success, even with a small library of compounds. In this article we have tried to give the reader an overview about sophisticated technologies and alternatives to peptides and nucleotides for protein purification and the art of ligand design for focused combinatorial synthesis.

Affinity Chromatography

The choice of purification strategy is largely contingent on its performance and economics, which is related to the effectiveness of the separation strategy employed and its robustness. Conventional purification techniques based on precipitation with salts, temperature, pH and high molecular weight polymers are now being substituted by highly selective and sophisticated strategies based on affinity chromatography. This highly selective technique simulates natural processes such as biorecognition. Molecular recognition encompasses interactions such as those between enzyme and substrates, antigens and antibodies, ligands and receptors, DNA and protein interactions, viral proteins and cell surface glycoproteins, hormones and transmitters that generate a cascade of events to carry out important biological activities. Exploitation of this affinity for interaction between target proteins and their complementary ligands or binding partners is the essence of all affinity techniques. The concept can be traced back to 1910 when Starkenstein first reported the isolation of α -amylase by adsorption onto insoluble starch, and was subsequently followed by several remarkable examples. Only in the second half of the 20th century did the use of the technique gain momentum and the technique was termed affinity chromatography in 1968 by Cuatrecasas. **Table 2** outlines the historical perspectives in the development of affinity chromatography. Ever since, techniques in affinity chromatography

Table 1 Quality concerns of purified recombinant protein products

Purity, efficacy, potency, stability, pharmacokinetics and pharmacodynamics
Toxicity and immunogenicity
Presence of contaminants such as nucleic acids, pyrogens, viruses, residual host cell proteins, cell culture media contaminants, leachates from separation media and unknown impurities
Post-translational modifications; mainly glycosylation and proteolytic processing
Protein folding and aggregation

Table 2 Important dates and achievements in the history of affinity techniques

1910	Starkenstein purified α -amylase on insoluble starch
1923	Lipase enrichment on powdered stearic acid
1951	Immunoaffinity chromatography
1967	Cyanogen bromide activation, <i>Staphylococcus</i> nuclease purified
1968	Term affinity chromatography coined by Cuatrecasas, biospecific adsorbents; a new era of modern affinity chromatography begins
1970	Group-specific adsorbents (coenzymes, lectins, nucleic acids)
1972	Boronates in affinity chromatography
1978	Textile dyes
1979	High performance liquid affinity chromatography (HPLAC)
1984	Biomimetic dyes
1985	Phage display
1986	Purification tags
1990	<i>De novo</i> ligand design

have been considerably refined for compliance with present-day strict quality control demands such as end product purity, safety, potency and stability. The choice of a stable and efficient solid support, activation and coupling chemistry and selection of a ligand have been investigated in detail since they contribute to efficient recoveries of the target protein.

The most commonly employed adsorbents in affinity chromatography are based on biomolecules such as monoclonal antibodies that offer selectivity and specificity. However, these adsorbents, besides being expensive, are prone to chemical and biological degradation, and themselves need purification prior to immobilization on a solid support and may have issues such as viruses and nucleic acids associated with them. Although biologicals have high selectivity, they have low capacities, a limited life cycle and a low scale-up potential. To counteract these problems, more durable and controllable peptides, peptidomimetics and synthetic ligands were introduced.

Peptide-based Ligands

In the last few years, peptide libraries have been a major area of development for the selection of biologically active peptides. These libraries offer the opportunity to study interactions between proteins and their natural ligand and are being used as potential drugs, antimicrobials and enzyme inhibitors, as bioactive peptides and as ligands for protein purification. Peptide libraries can be generated either by phage display or synthetic solid- or solution-phase chemical approaches. One of the most widespread and commonly used peptide-based technologies, phage display was conceived in 1985 by Smith and has been revolutionary in the synthesis, diversity and application of random peptide libraries in the search for novel protein-binding ligands for purification, structural and functional studies.

Random peptides are displayed on the surface of a filamentous bacteriophage (M13 or related bacteriophages) by fusion of the desired DNA sequence with either the minor (gIII) or major (gVIII) coat proteins of the bacteriophage. Multiple copies (five to thousands) of the peptide are then expressed on the bacteriophage surface owing to the presence of several copies of the coat proteins. Preparation of a large number of random oligonucleotides helps generate a combinatorial library of several millions of peptides. Selection of the phage particles expressing the peptide sequence with the desired activity and selectivity is performed through a process termed 'bio-panning'. The desired receptor for the peptides is immobilized on a solid support and the peptide-phage particles are loaded. The peptide-phage assemblies that do not bind wash off and the ones bound are eluted and amplified in *Escherichia coli*. The entire process is repeated 2–3 times to wash away any nonspecifically or weakly bound assemblies. The sequence of the selected peptide is determined by sequencing the coding region of the viral DNA.

The criteria for evaluation of peptide libraries includes generation of all possible combinations (millions) and numbers of peptide sequences to give the maximum probability of finding success. This aspect highlights the importance of rapid and efficient high-throughput screening for analysis of thousands or hundreds of millions of peptide candidates. The length of the sequences should be such that they include equimolar amounts of the tetramers and hexamers and the library should incorporate natural L-amino acids, their D-counterparts and unnatural amino acids.

In comparison with synthetic peptide synthesis, phage display is less labour-intensive and does not involve the painstaking synthesis of limited numbers of peptides. The peptide libraries can help localize

epitopes on the surfaces of antibodies (monoclonal or polyclonal), act as enzyme inhibitors, mimic cytokines, and DNA binding proteins may be used in the design of novel ligands for peptide receptors. The ability to identify regions in a protein that interact with certain peptides without dependence on structural data and other pre-existing data is remarkable. The opportunities in drug and ligand discovery are substantiated by a number of remarkable applications in the literature.

Oligonucleotide-based Ligands

Oligonucleotide (DNA or RNA)-based combinatorial chemistry has been exploited in a technology called SELEX (systemic evolution of ligand by exponential enrichment) to identify and isolate high affinity ligands directed for proteins, nucleic acids and low molecular weight targets such as peptides. The K_d values for such types of ligands are between 10^{-9} and 10^{-12} mol L⁻¹ for proteins and between 10^{-6} and 10^{-9} mol L⁻¹ for low molecular weight targets. The methodology involves random condensation of a mixture of activated monomers for generation of a combinatorial array of nucleic acid sequences that are assessed for binding towards any target (protein, nucleic acid, peptide or low molecular weight compound). The sequences are amplified by polymerase chain reaction, and the ones that show positive activity are selected and reamplified. The numbers of sequences that can be generated by commercially available synthesizers have been reported to be 10^{14} – 10^{15} in literature. The multiple rounds of selection exponentially enrich the group of oligonucleotide ligands with high affinity for the target.

Synthetic Ligands

Low molecular weight ligands that mimic the action of biologics have gained popularity not only in drug discovery and enzyme inhibition but also in several protein purification applications. The use of synthetic chemistry for the generation of compounds bound to a solid support (resin beads, silica surfaces or plastic pins) or in solution (for latter derivatization) is a fundamental yet powerful tool. However, nonspecific binding is a major disadvantage and has prompted the emergence of rationally designed ligands or biomimetics. These biomimetics have endearing properties such as high stability, defined chemical structure and toxicity, resistance to degradation, inexpensiveness and sterilizability *in situ*. They mimic the action of natural counterparts of proteins and can selectively extract the protein of interest. The devel-

opment of such synthetic molecules has been inspired and aided by the increasing availability of protein structural data and knowledge about protein–ligand complexes.

Synthetic combinatorial libraries comprising several hundreds to tens of millions of compounds are potential sources for drugs, enzyme inhibitors and antimicrobials. Peptidomimetic and organic libraries are yielding compounds to replace biomolecules for easier manipulation of the physical, chemical and biological properties and to address issues related to cost and availability. Oligomeric *N*-substituted glycines (NSG) or peptoids are novel polymers that are also being widely used for generation of vast chemical libraries. They are achiral, protease-resistant, inexpensive and are hydrolytically and enzymatically stable.

Design Rationale

A valuable approach in biotechnology involves the marriage of several technologies for the production of commercially viable diagnostics and therapeutics. The process of exploiting structural analysis, chemical synthesis and advanced computational tools when combined with rational design makes the technology more powerful, faster and logical. Structure elucidation of proteins by X-ray crystallographic and nuclear magnetic resonance (NMR) studies and improvement in software algorithms to generate homology models form the basis for rational ligand or drug design. Software programs provide us with an opportunity to calculate, visualize, formulate and hypothesize about the properties of molecules in terms of energy and orientation in their functionally active three-dimensional state and in complex with putative ligands.

The research strategy for the identification of key ligands typically involves obtaining structural information about the protein of interest, such as crystallographic, NMR or homology data. A target ligand-binding site is then identified on the protein; this may be an active site, a solvent-exposed region or motif on the protein surface or a site involved in binding the natural ligand. A combination of sophisticated modelling software packages and organic synthesis followed by activity analysis helps select a lead ligand. The optimum performance of the ligand is assessed by subjecting it to a range of experimental conditions and/or by constructing a library of near-neighbour ligands and selecting the one with desired key features. A general affinity (K_d) of 10^{-4} – 10^{-8} mol L⁻¹ between the immobilized ligand and the target protein proves satisfactory.

Design Tools

Modelling software can be used to probe properties of the three-dimensional structure of a protein such as charge distribution, key electrostatic interactions, hydrophobicity, hydrogen-bonding sites, extent of surface or solvent exposure of epitopes on the protein surface and identification of cavities, pockets and active sites. This information is essential for the site-directed *de novo* design of the ligand or drug candidates that may form novel leads. However, it is important to remember that proteins are elastic molecules that can bend and twist and even change their conformation upon ligand binding.

With an exponential increase in the availability of protein structures from the Brookhaven database (over 7000 protein entries), several sophisticated programs have emerged for modelling and visualization of molecules in three dimensions. SWISS-PROT, SCOP and the University College London (UCL) protein database (adapted from Protein Data Bank (PDB) files) are further sources for protein structures, with additional features such as LIGPLOTS of protein–ligand complexes. Homology models of proteins can be displayed with programs such as MODELLER that uses probability maps and COMPOSER that first defines the conserved regions, then the variations in structure such as loops and finally the side chains. Automated docking programs like DOCK and LUDI help predict the structure, mode and the binding free energy of the ligand–protein complexes. These programs work in conjunction with chemical structure databases such as the Cambridge Structural Database (Cambridge Crystallographic Data Centre with 110 000 compounds) or the Available Chemicals Database (100 000 compounds). HOOK is another program that generates putative ligands based on the chemical and steric complementarity between the ligand and the binding site on the protein. Calculation of molecular parameters is performed by force fields, with CHARMM, MM2, MM3 and AMBER being the commonly used programs. Force fields give information in terms of classical and mechanical potential energy functions rather than quantum mechanics.

The introduction of combinatorial synthesis has given us the capacity for multiple choices and alternatives. It is imperative not to be misled by the technique and synthesize a whole library of compounds that show no activity and cost money and time to synthesize. There are programs such as PRO-SELECT (SELECT = systematic elaboration of libraries enhanced by computational techniques) that help limit the size of a combinatorial library by using structural constraints on the protein target. Compact

molecular modelling software packages such as QUANTA, SYBYL, MACROMODEL and INSIGHT can be commercially obtained with a choice of energy minimization, autodocking and homology programs. Although there is a jungle of information and programs on drug and ligand design, these algorithms can only provide a certain rationale to rather serendipitous discoveries.

Designer Dyes

In 1968 Haeckel and others observed during gel filtration that Cibacron Blue F3G-A, the textile dye part of blue dextran, bound to pyruvate kinase. This observation subsequently led to the purification of a whole series of proteins and opened a new chapter in the role of textile dyes in protein purification. Dyes usually contain polyaromatic ring systems with electron withdrawing or donating groups and proteins contain a variety of hydrophobic and ionic residues. This complementarity assists in dye–protein interactions. Chlorotriazine dyes are inexpensive, easily synthesized and immobilized on solid support matrices and display a high capacity for proteins. The binding action mimics the binding of natural anionic heterocyclic substrates such as nucleotides, nucleic acids, adenosine triphosphate and coenzymes with enzymes. However, proteins (factor X, thrombin and kallikrein) that bind to cationic substrates have also been reported to bind nonspecifically to anionic dyes. Synthetic dyes have several advantages over biologics, although they can lack specificity. This prompted the *de novo* design and synthesis of biomimetic ligands aimed at specific sites on target proteins.

De Novo Ligand Design and Synthesis

Trypsin-like Family of Enzymes

The trypsin-like family of enzymes requires cationic substrates for their enzymatic action. Common examples are tissue plasminogen factor, factor Xa, thrombin kallikrein and urokinase. These enzymes are known to bind to the side chains of lysine or arginine at a site (primary binding pocket) proximal to the reactive Ser195. The specificity is introduced mainly by the side chain of Asp189 present at the bottom of the primary binding pocket. Secondary interactions formed by amino acids binding adjacent to the primary pocket also confer specificity to individual members of the enzyme family.

Tissue kallikrein, that has kininogen as the natural substrate, shows a marked preference for arginine in the primary binding pocket and phenylalanine on

the secondary site (Figure 1). The phenyl ring of phenylalanine is believed to form hydrophobic interactions with Trp215 and Tyr99. In contrast, trypsin demands less specificity from the secondary site residue. Thus, Burton and Lowe (1992) hypothesized that a structure based on the Phe-Arg dipeptidyl template should be designed and synthesized. The biomimetic ligand consisted of phenethylamine and *p*-aminobenamidine oriented on a triazine scaffold (Figure 2). Since the primary binding pocket of kallikrein lies in a depression, a hexamethylene spacer was introduced between the ligand and solid support to counteract any steric hindrance from the matrix backbone. On performing affinity chromatography with a crude pancreatic acetone powder, the designed synthetic ligand was able to purify kallikrein selectively with 110-fold purification factor. This efficiency compares well with the natural substrate ligands.

Artificial Protein A

Antibodies are used extensively in diagnostics, immunoassays, therapeutics and purifications, and can

be used as probes for labelling and imaging. Monoclonal antibodies, single chain and humanized antibodies have found innumerable applications in most areas of protein chemistry, biochemistry and molecular biology. However, the availability of antibodies in their highly pure form has largely limited their range of application. Immunoglobulins are routinely purified by immobilized staphylococcal protein A (SpA) or by conventional protein purification procedures. The high clinical value of immunoglobulins and disadvantages of using potentially toxic biologics in their purification initiated a remarkable study combining the powerful tools of rational computer-aided modelling and organic synthesis to generate an artificial protein A.

The crystal structure of the Fc domain of IgG and fragment B (Fb) of SpA (Figure 3) shows involvement of a total of 32 amino acid residues over an intersurface area corresponding to 400 nm². The primary forces holding the complex together are hydrophobic, hydrogen bonding and two salt bridges. The hydrophobic stacking is mainly provided by residues



Figure 1 (See Colour Plate 15) Illustration of the molecular model of porcine pancreatic kallikrein with the dipeptidyl motif Arg-Phe occurring in the natural kallikrein substrate, kininogen. The model was generated by manipulating the BPTI–pancreatic kallikrein complex using Quanta 97. The residues in the BPTI inhibitor not involved in the complex were deleted, leaving residues Lys-15 and Cys-14, which were substituted with arginine and phenylalanine respectively. The dipeptide was energy-minimized and its side chains were adjusted to interact with the primary and secondary binding sites, as the Lys-Cys dipeptide does in the BPTI–porcine pancreatic kallikrein complex.

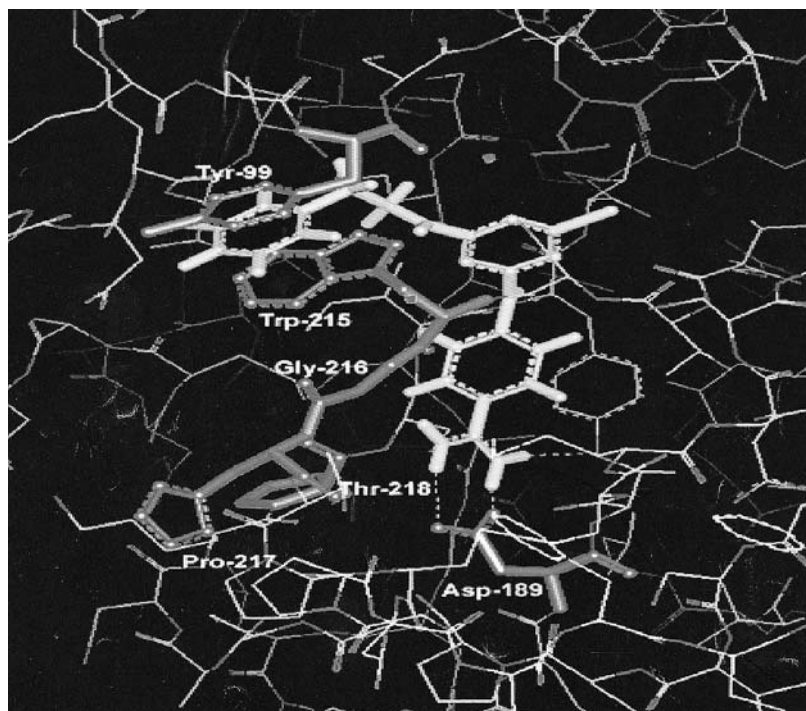


Figure 2 (See Colour Plate 16) Illustration of the synthetic ligand docked in the substrate-binding site of porcine pancreatic kallikrein. The ligand is an analogue of the Arg-Phe dipeptide that occurs in the natural substrate of kallikrein, kininogen, and is responsible for the enzyme-substrate complex. The benzamidine and phenethylamine moieties substituted on a triazine framework mimic the Arg-Phe dipeptide. The ligand was designed and energy-minimized in Quanta 97 and moved in the vicinity of the substrate-binding site, whence the side chains were adjusted to fit in the primary and secondary binding sites in porcine kallikrein. The aromatic ring of phenethylamine stacks in the primary binding site between Tyr-99 and Trp-215 and the benzamidine group forms several interactions with Asp-189, Ser-226, Gly-216, Pro-217 and Thr-218, forming the secondary binding site.

Phe124, Phe132, Tyr133, Leu136, Ile150 and the side chain of Lys154 on SpA and Ile 253 in IgG. Four hydrogen bonds can be identified between the ϵ^2 -amido group of Gln128 (SpA) and the γ -hydroxyl group of Ser254 (IgG), the δ^2 -amido group of Asn130 (SpA) and the δ^1 -carbonyl oxygen of Asn434 (IgG), the η -hydroxyl of Tyr133 (SpA) and the carbonyl oxygen of Leu432 (IgG), ϵ^2 -amido group of Gln311 (SpA) and the δ^1 -carbonyl oxygen of Asn147 (IgG). Salt bridges are formed between the δ -guanidino group of Arg146 (SpA) and the γ -carboxyl group of Asp315 (IgG) and between the ϵ -amino group of Lys154 and a sulfate ion in solution. The hydrophobic core dipeptide Phe132-Tyr133 on a helical twist of the SpA Fb region is oriented to interact with a shallow groove on IgG, comprising residues Leu251, Ile253, His310, Gln311, Glu430, Leu432, Asn434 and His435. This Phe-Tyr dipeptidyl motif is found in four highly conserved regions of SpA and each is capable of interacting with IgG from different species. If the binding pocket in IgG is made more hydrophilic by replacing the His435 with an Arg, the binding with SpA weakens, suggesting the importance of hydrophobic residues in the complex.

Li and co-workers (1998) noticed that this formed an exceptional basis for the design and synthesis of a ligand for the purification of IgG.

The biomimetic ligand (ApA) is comprised of anilino and tyramino substituents, mimicking the binding action of the Phe-Tyr dipeptidyl unit, spatially oriented on a triazine framework acting like the helical twist of SpA (Figure 4). A diethylamino spacer was used to immobilize the ligand on a solid support. This ligand proved to be complementary to the SpA binding site and had an affinity constant of 10^4 mol L^{-1} for IgG. ApA could purify 98% IgG from human plasma and also inhibit the binding between SpA and IgG on enzyme-linked immunosorbent assay. Immobilized ApA showed a capacity of 20 mg IgG per gram moist weight gel. This biomimetic ligand could be further optimized to increase selectivity, capacity and the use of milder experimental conditions. Thus, a combinatorial library comprising 88 adsorbents was constructed for lead optimization of ligand ApA. The synthesis was inspired by the 'mix and split' procedure on a triazine scaffold and was intended to mimic the binding mechanism of ApA with improved features.

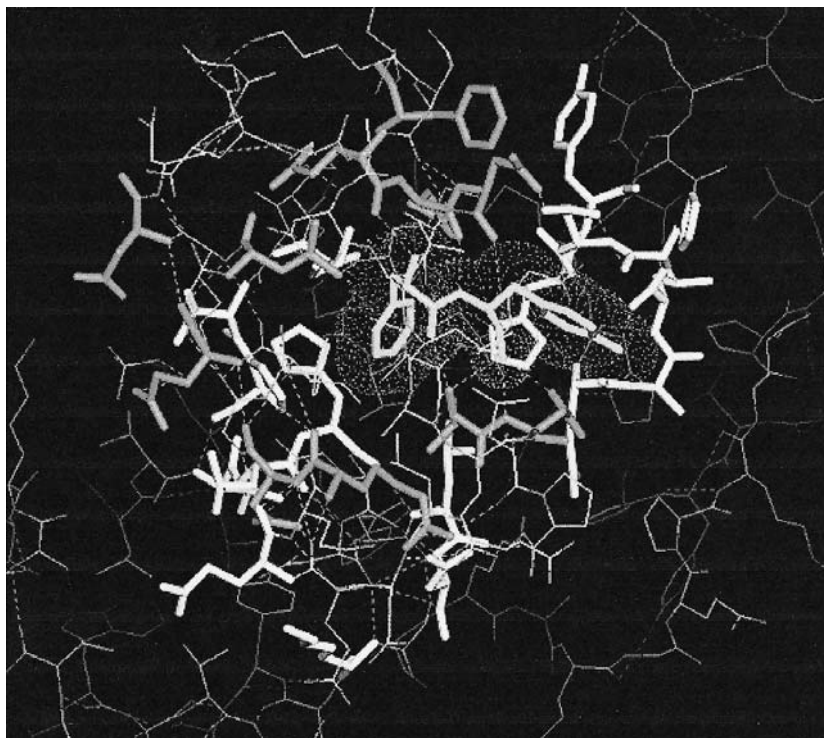


Figure 3 (See Colour Plate 17) The complex between the Fb fragment of SpA and the Fc fragment of IgG. The residues in pink represent amino acids in SpA interacting with the residues in yellow in IgG. The residues in green represent the Phe-Tyr dipeptide, key residues in holding the complex. The dotted lines represent inter- and intramolecular hydrogen bonding. The interaction involves a total of 32 amino acids spanning an intersurface area corresponding to 40 nm². The interaction is predominantly characterized by hydrophobic interactions as well as some hydrogen bonding and two salt bridges.

Combinatorial solid-phase synthesis and activity analysis suggested that a biomimetic (22/8) with 3-aminophenol and 4-amino-1-naphthol substituted on a triazinyl scaffold was able to purify IgG selectively from diluted human plasma and eluted more than 99% of bound IgG.

A similar approach was employed to design highly specific ligands for other industrially and clinically important recombinant proteins such as factor VIII and insulin. These designer ligands proved to be highly successful and were able to isolate target proteins from crude fermentation broth with high specific activities and purification factor.

Protein Structural Isoforms

Protein synthesis is accompanied by some important events, termed post-translational modifications, that convert the information carried by a two-dimensional polypeptide into a complex, biologically active three-dimensional protein. These modifications encompass events such as protein folding and phosphorylation, glycosylation, sulfation and myristoylation. Variations in post-translational modifications lead to variants of a protein that are products of the same

gene but may vary in structure, function, pharmacokinetics and pharmacodynamics. Glycosylation, the most important event during protein synthesis, has received much attention in the last few years due to its influence on the clinical and therapeutic properties of proteins. Most biopharmaceuticals produced by recombinant DNA technology, such as erythropoietin and tissue-plasminogen activator for *in vivo* administration, are glycosylated. These may suffer from glycoform heterogeneity due to variations in the carbohydrate sequence or due to variable site occupancy of the sugar moieties on the protein. Consequently, it is important to isolate, resolve and analyse recombinant glycoforms with defined glycosylation and biological properties prior to clinical prescription.

The strategy for the resolution of glycoforms involves generation of synthetic ligands that display affinity and selectivity for the sugar moieties on glycoproteins but have no interaction with the protein *per se*. A detailed assessment of native protein-carbohydrate interactions using molecular modelling tools formed the basis for the synthesis of carbohydrate-binding ligands. The ligands were synthesized on a solid phase and assessed for their sugar-binding ability with glycoenzymes. Chromatography

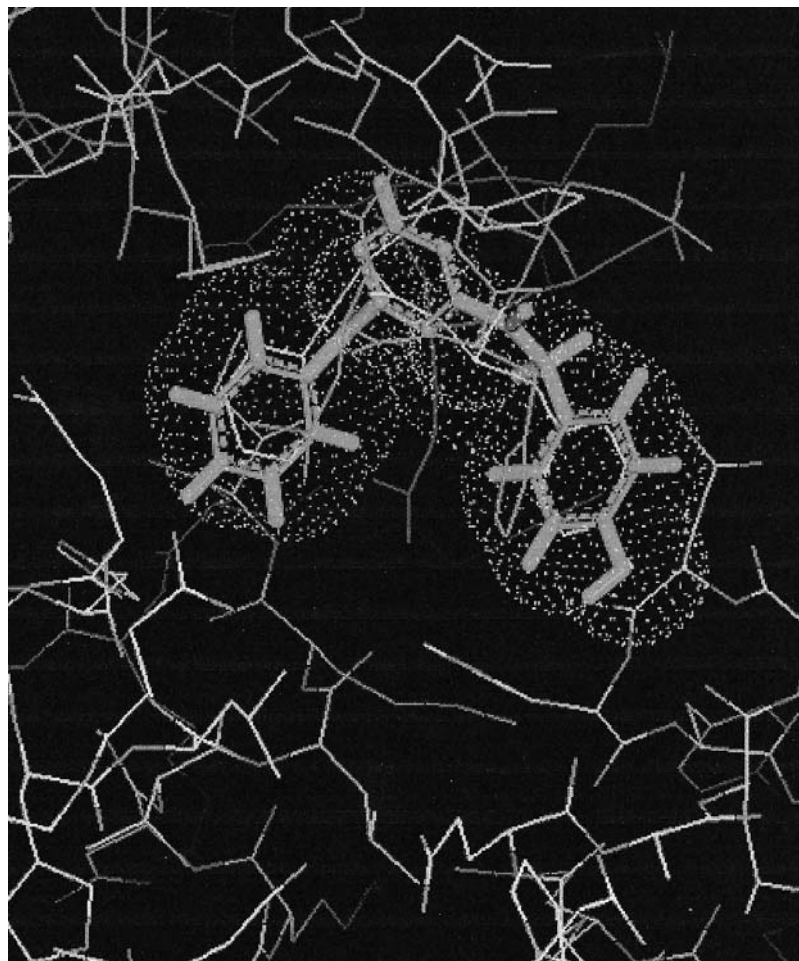


Figure 4 (See Colour Plate 18) Molecular model between the Fc fragment of IgG and the synthetic ligand ApA. The ligand, in red, is a mimic of the key dipeptide Phe-Tyr, in green, in SpA and comprises anilino and tyramino moieties substituted on a triazinyl framework. ApA is located at the putative binding site among the amino acid residues involved in the interaction between the Fc part of IgG and Fb fragment of SpA.

with partially and completely deglycosylated enzymes, elution of glycoproteins with sugars and inhibition of glycoproteins for the synthetic receptor in the presence of free sugars helped to assess the selectivity of the ligands for the sugars. Palanisamy and co-workers (1999) observed that triazine-based ligands display selectivity for the glycomoiety on glycoproteins and identified mannose-binding synthetic ligands.

Applications

Readers can put their imagination to the test and list innumerable applications for this technology. Reliable and robust synthetic affinity ligands are being used for the purification of several therapeutically and industrially important recombinant proteins for clinical, diagnostic or research purposes. The opportunity to use efficient, nontoxic and highly characterized synthetic compounds is lucrative not only in protein purification but also in localization

(probes) and tagging of macromolecules. A major application is in the removal of contaminants such as endotoxins, misfolded protein conformations, nucleic acids and viruses during recombinant protein production.

The applications of these synthetic low molecular weight ligands as drugs are forthcoming and these highly specific, nontoxic, characterized, inexpensive, easy-to-synthesize and nonlabile compounds could prove to be blockbuster therapeutics. The direct use of these ligands as drugs in diseases such as cancer stems from the fact that during infection there is an increased vasculature for nutrient supply. This is aided by overexpression of certain proteins that, if blocked from interacting with their complementary receptors, could help circumvent tumours.

Application of peptides and oligonucleotides as drugs, high affinity ligands or potent inhibitors is still in its infancy. The problems of low oral activities and susceptibility to hydrolysis has initiated tagging

the products with conjugates such lipophilic agents and polymers that not only reduce degradation but may enable efficient uptake by cells. However, a putative application is in identifying epitopes on proteins for receptor-binding studies, drug design and inhibition of proteases, proteins and oncogenes. In fact, the technology could complement rational ligand design when structural data for proteins are absent.

Bioprocess monitoring is another application that stems from requirements for optimizing cultivation conditions and harvesting times for recombinant protein products in fermentors. Modern bioprocessing demands control of several processing parameters with engineering ingenuity to obtain online monitoring to reduce labour, cost and analysis time. The ability to measure specific macromolecules in a fermentor in the presence of a soup of nutrients, host cell proteins, media components and other impurities requires highly selective and efficient ligands that can also be sterilized *in situ*. The selective biorecognition system, when incorporated with a sensitive transducer for interpreting the biorecognition signal into a comprehensible electric signal, generates a reliable biosensor for online measurements.

Conclusions and Future Prospects

Success in biotechnology and biochemistry is well attributed to developments in analytical techniques and instrumentation. The production of several high value therapeutics would not have gone to completion if it were not for efficient methods in protein purification. Conventional purification techniques are becoming outdated for industrial applications, and technologies based on proteins, peptides and oligonucleotides have gained an impetus, although they are prone to enzymatic and chemical degradation and have short *in vivo* and *in vitro* lifetimes. Peptides and oligonucleotides make undesirable drug candidates owing to poor oral activities. Furthermore, peptides have a large number of conformations that may be highly flexible and which can make attainment of the most favourable conformation and orientation difficult and lead to low affinity ligand-protein complexes.

The current focus is on compounds that lack the repetitive backbone units linked by facile bonds prone to chemical and biological (proteases and nucleases) cleavage. Thus, large number of alternative methodologies have been proposed to replace the peptide bond (peptidomimetics) with the use of peptoids, carbamates, sulfones, alkenes, urea, phosphodiesteres and sulfonamides.

Sophisticated organic synthesis has inspired chemists to synthesize receptors, drugs, peptides, nucleo-

tides and their mimetics to identify epitopes on surfaces of proteins for therapeutic, diagnostic, inhibitory and purification purposes. A combinatorial approach has become the choice of several chemists to maximize the possibility of success for lead generation, optimization and development of candidate drugs or ligands. However, it is very expensive and laborious to produce, screen and manage data for all these compounds. Some direction and rationale is essential. Information about structure-activity relationship lends that extra information that can help in pre-selection.

The availability of fast and sophisticated computer modelling software packages and the marriage of rationale and serendipity makes the design and synthesis a successful venture for investigation. A prerequisite in *de novo* design of biomimetics ligands is the availability of protein structural data and information on protein-ligand complexes. With the advances in proteomics, protein structural availability is hardly a limiting issue. In the coming years, there will be an explosion of protein structural data. The completion of the Human Genome Project will unveil countless new proteins, challenging methodologies such as random screening, rational design, combinatorial chemistry and high-throughput screening to become rapid, inexpensive and fully automated with the incorporation of robotics. Simultaneous developments in data management, proteomics and bioinformatics will integrate the technology and reduce time, cost and labour. The challenge to mimic the action of natural biological recognition is ongoing, and, as nature uses only a repertoire of 20 amino acids to generate an endless combinatorial list of proteins, the challenge for rational ligand design continues.

See Colour Plates 15, 16, 17, 18.

See also: II/Affinity Separation: Affinity Membranes; Biochemical Engineering Aspects; Covalent Chromatography; Dye Ligands; Imprint Polymers; Theory and Development of Affinity Chromatography. **Appendix 2/Essential Guides to Method Development in Affinity Chromatography.**

Further Reading

- Burton NP and Lowe CR (1992) Design of novel affinity adsorbents for the purification of trypsin-like proteases. *Journal of Molecular Recognition* 5: 55.
- Cuatrecasas P and Anfinsen CB (1971) Affinity chromatography. *Annual Reviews in Biochemistry* 40: 259.
- Haeckel R, Hess B, Lauterborn W and Wuster KH (1968) Purification and allosteric properties of yeast pyruvate kinase. *Hoppe Seylers Zeitschrift für Physiological Chemistry* 349(5): 699
- Li R, Dowd V, Stewart DJ *et al.* (1998) Design, synthesis, and application of a protein A mimetic. *Nature Biotechnology* 16: 190.

- Lowe CR, Burton SJ, Burton NP *et al.* (1992) Designer dyes: 'biomimetic' ligands for the purification of pharmaceutical proteins by affinity chromatography. *TIBTECH* 10: 442.
- Palanisamy UD, Hussain A, Iqbal S *et al.* (1999) Design, synthesis and characterisation of affinity ligands for glycoproteins. *Journal of Molecular Recognition* 12: 57.
- Smith GP (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science* 228: 1315.
- Starkenstein EV (1910) Über fermentenwirkung und deren beeinflussung durch neutralisalz. *Biochemische Zeitschrift* 24: 14.
- Teng SF, Sproule K, Hussain A and Lowe CR (1999) A strategy for the generation of biomimetic ligands for affinity chromatography. Combinatorial; synthesis and biological evaluation of an IgG binding ligand. *Journal of Molecular Recognition* 12: 67.
- Tuerk C and Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249: 505.

Theory and Development of Affinity Chromatography

R. Scopes, La Trobe University, Bundoora, Melbourne, Australia

Copyright © 2000 Academic Press

Introduction

The term affinity chromatography began to be used extensively in the 1960s to describe protein separation methods that made use of the specific biological interaction of the desired protein with some ligand that was immobilized on an adsorbent matrix. Since most proteins, and all enzymes, bind to some compound very specifically, this immediately promised to solve most protein purification problems. But, as with all good ideas, there were many cases when it did not work as expected; the general concept of affinity chromatography for purifying proteins found its niche, but was no panacea. More recently, it has found a fairly widespread application in purifying recombinant proteins, using very standardized procedures.

Although most applications have been for proteins, it is not so limited in theory, since other biological macromolecules have specific interactions which can be exploited, especially nucleic acids. But, for the purposes of this article, the principles will be expounded with proteins as the prime target. We should consider the definition(s) of affinity chromatography carefully, since it does not mean the same to everyone. First, the word affinity. Any two components that are attracted to each other can be said to have an affinity, but if we took that as a definition, the term would be too broad to be useful – for instance, it could include all types of chromatography. It is better to limit the definition of affinity to a biologically significant interaction such as between a hormone and its receptor, an enzyme and its substrate, or an antibody and its antigen. Unfortunately, there are well-established uses of the term, such as immobilized metal affinity chromatography, in which

the interaction is not biologically relevant, though it too can be highly specific. Perhaps a better definition could imply simply a high specificity and selectivity of the interaction, though that can exclude some examples of true biological affinity.

The other word, chromatography, strictly means that process in which components are adsorbed and desorbed continuously as they move down a column, or through some other medium, resulting in a multi-stage separation of different components according to their partitioning between the stationary and mobile phases. But affinity methods are usually treated in an 'on-off' fashion in which, after total adsorption of the desired component, a stepwise change in the buffer mobile phase results in its complete elution, and true chromatography is not carried out. Nevertheless, the word chromatography is used more widely than its strict definition, to encompass any use of an adsorbent, even in this 'on-off' fashion.

And so we come up with a definition of affinity chromatography as a chromatographic procedure utilizing an adsorbent involving an immobilized ligand which has a high specificity for binding the desired component, preferably to the exclusion of all others. This binding can be loosened by a change in buffer conditions, to elute the desired component relatively free of contaminants. If the ligand is the natural biological ligand of the desired component, then the more precise term bioaffinity chromatography can be used. On the other hand, when the interaction is specific, but the ligand is unnatural, terms such as pseudo-affinity chromatography and biomimetic chromatography have been adopted.

Early Developments in Affinity Chromatography

The first protein to be purified by affinity chromatography was amylase in 1910, for which a column